Hierarchical Methods of Moments

Matteo Ruffini [1] Guillaume Rabusseau [2] Borja Balle [3]

[1]Universitat Politècnica de Catalunya

[2]McGill University

[3]Amazon Research Cambridge

Background: methods of moments

Methods of moments:
a unified framework to learn latent variable models.

Task: From iid samples $\rightarrow$ model parameters $(\mu_1, ..., \mu_k, \omega)$

Background: methods of moments

Methods of moments:
a unified framework to learn latent variable models.

Task: From iid samples $\rightarrow$ model parameters $(\mu_1, ..., \mu_k, \omega)$

1. Estimate the moments:

$$M_1 = \sum_i \omega_i \mu_i \in \mathbb{R}^d$$

$$M_2 = \sum_i \omega_i \mu_i \otimes \mu_i \in \mathbb{R}^{d \times d}$$

$$M_3 = \sum_i \omega_i \mu_i \otimes \mu_i \otimes \mu_i \in \mathbb{R}^{d \times d \times d}$$

2. Parameters via tensor decomposition:

$$\mathcal{TD}(M_1, M_2, M_3, k) \rightarrow (\mu_i, ..., \mu_k, \omega)$$

Background: methods of moments

Methods of moments:
a unified framework to learn latent variable models.

Task: From iid samples $\rightarrow$ model parameters $(\mu_1, ..., \mu_k, \omega)$

1. Estimate the moments:

$$M_1 = \sum_i \omega_i \mu_i \in \mathbb{R}^d$$

$$M_2 = \sum_i \omega_i \mu_i \otimes \mu_i \in \mathbb{R}^{d \times d}$$

$$M_3 = \sum_i \omega_i \mu_i \otimes \mu_i \otimes \mu_i \in \mathbb{R}^{d \times d \times d}$$

2. Parameters via tensor decomposition:

$$\mathcal{TD}(M_1, M_2, M_3, k) \rightarrow (\mu_i, ..., \mu_k, \omega)$$

Examples:

- Single Topic Model (μ_i: topics)
- Mixture models (μ_i: centers)
- HMM (μ_i: emissions probs.)
- Many others..

Motivation of this paper

Methods of moments are appealing:

- Run in polynomial time.
- Single pass through the data.
- Provable guarantees of optimality **in the realizable setting**.

Out of the model are often unstable.

Motivation of this paper

Methods of moments are appealing:

- Run in polynomial time.
- Single pass through the data.
- Provable guarantees of optimality **in the realizable setting**.

Out of the model are often unstable.

But in the real world..

- Data never comes from a specific model.
- Unknown/too big number of latent states.

SIDIWO: a new method of moments

SIDIWO:

Simultaneous **Di**agonalization based on **W**hitening and **O**ptimization.

SIDIWO: a new method of moments

SIDIWO:

Simultaneous **Di**agonalization based on **W**hitening and **O**ptimization.

$$(x^{(1)}, ..., x^{(n)}) \underbrace{\rightarrow}_{Moments} M_1, M_2, M_3$$

$$SIDIWO(M_1, M_2, M_3, l) \rightarrow \tilde{\mu}_i, ..., \tilde{\mu}_l, \tilde{\omega}$$

SIDIWO:

Simultaneous **Di**agonalization based on **W**hitening and **O**ptimization.

$$(x^{(1)}, \dots, x^{(n)}) \underbrace{\rightarrow}_{Moments} M_1, M_2, M_3$$

$$SIDIWO(M_1, M_2, M_3, l) \rightarrow \tilde{\mu}_i, \dots, \tilde{\mu}_l, \tilde{\omega}$$

The optimal l-states model approximating the model generating the data.

$\tilde{\mu}_1, \dots, \tilde{\mu}_l \in \mathbb{R}^d$: **pseudocenters.**

Each pseudocenter synthetically represents some true centers.

Application: hierarchical method of moment

The pseudocenters approximate the original centers of the model.

$$\Downarrow$$

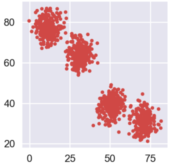
Hierarchical learning of latent variable models.

Application: hierarchical method of moment

The pseudocenters approximate the original centers of the model.

$$\Downarrow$$

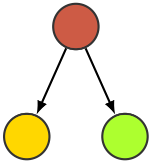
Hierarchical learning of latent variable models.

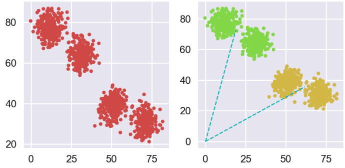


The pseudocenters approximate the original centers of the model.

$$\Downarrow$$

Hierarchical learning of latent variable models.



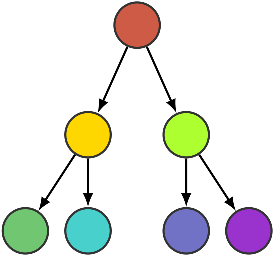


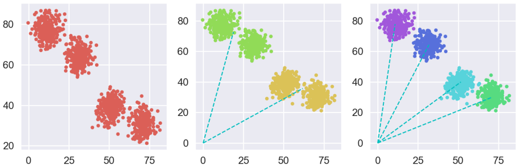
Application: hierarchical method of moment

The pseudocenters approximate the original centers of the model.

$$\Downarrow$$

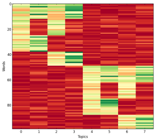
Hierarchical learning of latent variable models.

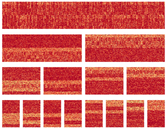




Experiments

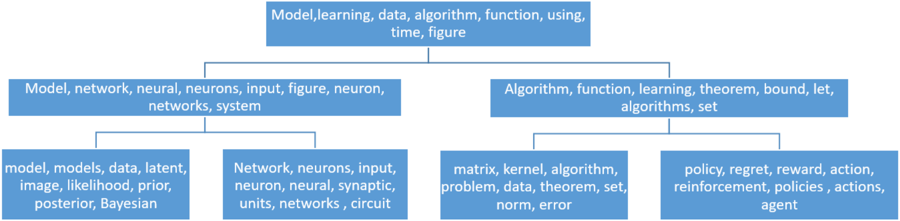
Synthetic Data





Method	Adj. Rand Idx Mean	St. dev.	Run. Time
TPM	0.93	0.06	1.2 sec.
SVD	0.52	0.13	0.1 sec.
Rand. Proj.	0.72	0.06	16 min.
SIDIWO	0.98	0.01	0.4 sec.

Real Data: NIPS papers 1987-2015



See you at the conference!

We are at poster #49 @Pacific Ballroom